

# Evaluating the Translation of Figurative Language: A Comparative Study of ChatGPT and Human Translators

**Yousef Sahari**

Department of English Language and Literature, College of Arts and Letters, The University of Bisha,

Bisha, Saudi Arabia, Email: [ysahari@ub.edu.sa](mailto:ysahari@ub.edu.sa)

<https://orcid.org/0000-0001-8318-6987>

**Fawaz Ali Ahmed Qasem**

Department of English language and Literature, College of Arts and Letters, University of Bisha,

Bisha, Saudi Arabia, Email: [faqasem@ub.edu.sa](mailto:faqasem@ub.edu.sa)

<https://orcid.org/0000-0002-2314-1768>

**Eisa Ahmad S. Asiri**

English Department, King Khalid University, Abha, Saudi Arabia, Email: [eaieisa@kku.edu.sa](mailto:eaieisa@kku.edu.sa)

<https://orcid.org/0000-0002-5498-4389>

**Ibrahim Alasmri**

English Department, King Khalid University, Abha, Saudi Arabia, Email: [Ifasmary@kku.edu.sa](mailto:Ifasmary@kku.edu.sa)

<https://orcid.org/0000-0003-2323-3176>

**Ahmad Assiri**

Department of General Courses, King Abdulaziz University, Jeddah, Saudi Arabia, Email:

[aassiri1@kau.edu.sa](mailto:aassiri1@kau.edu.sa)

<https://orcid.org/0009-0004-5800-3930>

**Shafi Saad Alqahtani**

Department of English Language and Literature, College of Arts and Letters, University of Bisha,

Saudi Arabia

[shafy@ub.edu.sa](mailto:shafy@ub.edu.sa)

**Hassan Saleh Mahdi**

Faculty of Language Studies, Arab Open University, Riyadh, Saudi Arabia

Email: [almatari7@gmail.com](mailto:almatari7@gmail.com)

ORCID ID: <https://orcid.org/0000-0003-4023-8219>

## Abstract

Generative Pretrained Transformer (GPT) has demonstrated significant advancements recently across various language tasks, including machine translation. However, many studies assessing ChatGPT's efficacy in translation tasks have primarily focused on generic texts. Therefore, this study aimed to evaluate the performance of GPT-3 in translating figurative language between Arabic and English, comparing it against human translators. The study utilized several passages containing figurative language on diverse topics to achieve this objective. Both GPT-3 and a proficient human translator translated these passages, and qualitative criteria were employed to assess their translation performance, including accuracy, fluency, coherence, cohesion, and handling of metaphorical language. The findings indicated that, unlike human translation, GPT-

3 generally produced translations that were comprehensible but struggled to fully capture the differences of figurative language. Specifically, when translating from English source texts (STs), GPT-3 generally maintained a similar number of sentences. In contrast, for Arabic STs, GPT-3 tended to break down and shorten lengthy sentences. The study suggests that while human translators excel in complex translations requiring cultural and idiomatic insights, AI can effectively handle simpler tasks. This highlights the potential for hybrid translation models that leverage AI efficiency alongside human expertise.

**Keywords:** AI-based translation, ChatGPT, figurative language, human translation, translation.

## Introduction

Programs in artificial intelligence have gained significant traction in academia, offering researchers and students exposure to diverse and innovative areas of knowledge and experience (Crompton & Burke, 2023). One such AI system, ChatGPT, has been specifically trained for text generation, predicting subsequent words within a sequence based on algorithmic patterns of word coherence (Cingillioglu, 2023). ChatGPT exhibits exceptional capabilities, including proficiently authoring scholarly articles (Eke, 2023). It excels in tasks such as summarization, text expansion, and adaptation to present multiple perspectives swiftly and flexibly. As an AI tool poised for academic translation, further exploration of ChatGPT across various languages and contexts is warranted (Javaid et al., 2023). Figurative language translation represents a significant area of investigation in this regard.

Translating figurative language poses one of the greatest challenges for translators across different texts and genres. This difficulty arises from the use of intricate vocabulary, alongside a variety of literary devices such as metaphor, simile, and culturally specific imagery (Abdel-Raheem, 2024). Figurative language, characterized by vivid and concrete expressions, is commonly employed in narratives and descriptive texts across genres (Qiong & Xiaobing, 2005). Therefore, this study aims to evaluate ChatGPT's capacity and effectiveness in translating figurative language through a corpus analysis of translational texts, comparing its performance with human translations.

This paper employs two translation methodologies—human translation and GPT-3 translation—to explore the differences of translating figurative language, including idioms, metaphors, and similes, between Arabic and English. Despite significant strides made by ChatGPT and other AI translation technologies in rendering plain text (Lee, 2023), their proficiency in translating metaphorical text remains contentious. This is crucial as metaphorical language plays a pivotal role in preserving the emotional and cultural depth of the original text. Additionally, the study delves into the ongoing debate surrounding whether artificial intelligence can or should replace human translators. The advancement of AI technology has sparked discussions on the potential superiority of machines in translation tasks, raising concerns about job security within the translation industry and the overall quality of machine-generated translations.

## Literature review

### *Figurative language*

According to Montgomery et al. (2007), figurative language involves the use of words or phrases to express ideas beyond their literal meaning, aiming for a more precise and potentially more accurate understanding. Nida (1975) further elaborates that figurative words serve as substitutes for alternative meanings or expressions, connected by ideas rather than exact synonyms, often conveyed through idioms, similes, and metaphors. Abdul Wahid (2017) underscores the immense difficulty in translating figurative language, emphasizing the need for fluency in both source and target languages, meticulous word choice, and adept translation techniques. One of the primary challenges lies in effectively conveying the cultural differences embedded within the original text.

Bell and Candlin (2016) acknowledge that inherent differences in languages, including distinct codes and grammatical norms, hinder achieving complete equivalence in translation. Consequently, translating figurative language entails more than merely substituting terms with their counterparts in another language. Specifically, translating figurative language from Arabic to English demands a profound understanding of the cultural contexts and figurative elements inherent to both languages, often necessitating different rephrasing.

Baker (2018) identifies four main challenges in translating metaphorical language: the risk of literal translations from the source language becoming idiomatic in the target language; the absence of direct equivalents for figurative expressions in the target language; the existence of equivalent

figurative expressions whose interpretations vary depending on context; and disparities in the discourse, context, and frequency of figurative language use between source and target languages. To tackle these challenges, Baker (2018) proposes two translation approaches: (i) utilizing figurative expressions with comparable meanings and forms, and (ii) employing figurative language with comparable meanings but different forms, achieved through paraphrasing. According to Baker (2018), paraphrasing becomes necessary when an exact counterpart does not exist in the target language.

### *ChatGPT and translation*

ChatGPT, short for Generative Pre-trained Transformer, represents a cutting-edge language model engineered to comprehend and generate text akin to human language. Through extensive training on diverse datasets, ChatGPT excels in numerous language-related tasks such as answering questions, creating original content, and aiding in natural language understanding (Javaid et al., 2023). It has showcased its capability to interpret and respond to text inputs in a manner closely resembling human conversation, leveraging its comprehensive training (Cai et al., 2023). One significant application of ChatGPT lies in translation. Utilizing its grasp of context and semantics, ChatGPT facilitates smooth translations across different languages. Users can input text in one language, and ChatGPT produces a coherent translation in the desired language, proving invaluable for individuals and businesses engaged in multilingual communication. Central to ChatGPT's effectiveness in translation is its ability to understand context, ensuring accurate and contextually relevant translations by considering surrounding words and phrases (Aghai, 2024; Lee, 2023). Unlike conventional translation tools that may specialize in specific language pairs, ChatGPT's training encompasses a wide array of languages. This multilingual proficiency enables it to handle translations across numerous language combinations, offering flexibility for users dealing with diverse linguistic contexts. Furthermore, ChatGPT can adapt its responses to match the user's style, enhancing the naturalness and alignment of translated content with the user's preferred tone and communication style (Al-Hasan et al., 2024).

ChatGPT has demonstrated the capability to generate abstracts that closely mimic human-created content, as shown by Gao et al. (2022). Several studies have investigated ChatGPT's translation capabilities. For example, Jiao et al. (2023) assessed ChatGPT's translation performance by examining factors such as translation prompts, multilingual translation, and translation robustness. Their research revealed that while GPT-3 performed well compared to commercial translation tools like Google Translate for high-resource European languages, it had notable limitations for lower-resource or more distantly related languages. However, the introduction of the GPT-4 engine significantly enhanced ChatGPT's translation performance, making it competitive with commercial products even for distant languages.

Additionally, Banat and Adla (2023) examined GPT-3's effectiveness in translating specialized Arabic texts into English, comparing its performance to that of human translators. Their study involved translating ten chapters from a specialized Arabic book on a specific religious context, with translations produced by both a professional human translator and GPT-3. Qualitative measures were used to evaluate GPT-3's translation against the human version. The results indicated that GPT-3 produced generally understandable translations but struggled with cultural context differences. Nonetheless, GPT-3 achieved a high level of accuracy in translating specialized religious texts, with scores comparable to human translations in some cases.

Hendy et al. (2023) explored GPT models in the context of machine translation, assessing various aspects including the quality of different GPT models relative to state-of-the-art research and commercial systems, the impact of prompting strategies, robustness against domain shifts, and document-level translation. Their findings revealed that GPT models delivered highly competitive translation quality for well-resourced languages but faced limitations with lower-resourced languages. The study also highlighted that hybrid approaches, combining GPT models with other translation systems, effectively enhanced translation quality. Furthermore, Khoshafah (2023) evaluated GPT-3's translation abilities by comparing its outputs to those of human professionals across diverse genres, including history, literature, media, legal documents, and scientific articles. The study found that GPT-3 was effective for translating basic content, but its performance declined with complex texts requiring human expertise.

### *Research aims*

Recent advancements in AI, particularly with models like Chat GPT, have shown significant promise in translating various genres of writing, from technical documents to literary works. These AI tools have been increasingly employed to handle a wide range of linguistic challenges, offering translations

that often rival those produced by human experts. However, despite these advancements, there remains a notable gap in research regarding the effectiveness of AI in translating figurative language. Figurative language, with its distinct and often context-dependent meanings, poses a unique challenge that has not been extensively studied in the context of AI translation tools. To address this gap, this research aims to evaluate the effectiveness of Chat GPT in translating figurative language compared to human translators. The study is an attempt to answer the following question.

*RQ1. How effective is GPT-3 in translating figurative language compared to human translation?*

## **Method**

This study aims to explore the effectiveness of the AI tool GPT-3 in translating figurative language. Chat GPT-3 was chosen for this study due to its free accessibility. The comparison between Chat GPT-3 and GPT-4 is beyond the scope of this research, as both models can generate similar outputs. To achieve the study's objectives, four texts were selected and translated by both a human professional translator and GPT-3. The quality of GPT-3's translations was evaluated using qualitative measures, specifically by comparing them to human translations through direct assessment. Human judges rated the translations based on factors such as fluency, accuracy, and naturalness. The human raters involved were three university-level instructors.

This approach allowed for a detailed understanding of GPT-3's translation strengths and weaknesses and identified areas where human translators still outperform machine translation systems. The evaluation criteria, based on those proposed by Banat and Adla (2023), included accuracy, fluency, cohesion, and coherence. An additional criterion, translating figurative language, was added for this study. Accuracy refers to how well the translation conveys the original text's intended meaning, while fluency measures the naturalness of the translation in the target language. Cohesion and coherence assess the organization and connectedness of the translation.

Furthermore, a quantitative analysis was conducted for each text, comparing the number of sentences in the source text (ST) and translated texts, the number of words in the ST and translated texts, and the average sentence length in the three versions of the texts. These analyses were performed using specialized websites designed to measure these aspects. By employing these criteria and analyses, the study aimed to gain insights into the strengths and weaknesses of GPT-3's machine translation and identify areas for improvement.

### *Data analysis procedures*

In this study, the data analysis procedures encompassed both quantitative and descriptive analysis to evaluate translations from human translator and GPT-3. The quantitative analysis involved measuring the number of sentences, the number of words, and the average sentence length across the source text, human translations, and GPT-3's translations. Descriptive analysis provided a deeper examination of accuracy, fluency, cohesion, and coherence, utilizing specific examples to illustrate these aspects.

## **Result**

To address the research question, this study utilized four texts: two written in Arabic and two in English. These texts were translated by both a human translator and GPT-3. They underwent both quantitative and qualitative analysis. The quantitative analysis examined the number of sentences and words in the source texts and their translations, as well as the average sentence length in both the original and translated texts. The subsequent section presents the analysis of these texts.

### *Text 1*

Source text      On the other hand, old age is an expensive school because it is a storehouse of human experience. The truth is that experience charges high school fees but it is better than any other school. An old man can be of great benefit to the community through his personal experience which he could formulate into literary works, biographies or sketches of art or interesting talks. Probably, the experience of old people and the vivacity of young people would be the best combination. I do not think that such a combination would be impossible. It depends on the old man himself who could imagine life as a spring of merriness and pleasure and listen to nature calling us in numberless songs and feel that the banquet of nature is the banquet of love itself and that life void of love would be trivial for the young and the old alike. Then an old man would be able to invest the years of his life for enjoying

life. Accordingly, old age might be synonymous with youth in such a way that people would look upon it as a form of progressive vividness.

Human translation	من جهة ثانية فالشيخوخة مدرسة غالية لأنها مستودع للخبرة البشرية، والحق أن الخبرة تتقاضى أجوراً مدرسية عالية، لكنها أفضل من أية مدرسة أخرى. فالشيخ يستطيع أن يقدم إلى المجتمع نفعاً كبيراً من خلال تجاربه الشخصية التي يبليها أحياناً أعمالاً أدبية أو سيراً ذاتية أو لوحات فنية أو أحاديث شيقة. لعل أفضل مزيج يتمثل في خبرة الشيوخ وحيوية الشباب، ولا أحسب أن مثل هذا الجمع مستحيل، فالأمر يعتمد على الشيخ نفسه إذا استطاع أن يتصور الحياة ينبوع مرح ومسرة، وتمكن من أن يصغي إلى الطبيعة التي تتأدنا بأغان لا عدد لها، وشعر أن وليمة الطبيعة هي وليمة الحب ذاته، وأن الحياة الخالية من الحب تافهة بالنسبة إلى الشباب وإلى الشيوخ على حد سواء. عندها يصبح الشيخ قادراً على استثمار سنوات حياته من أجل الاستمتاع بالحياة. وعلى هذا النحو يمكن أن تكون الشيخوخة مرادفة للشباب بحيث ينظر الناس إليها على أنها شكل من الحيوية المستمرة.
GPT-3's translation	من ناحية أخرى، الشيخوخة هي مدرسة مكلفة لأنها مخزن لتجربة الإنسان. الحقيقة هي أن التجربة تفرض رسوماً مدرسية باهظة، لكنها أفضل من أي مدرسة أخرى. يمكن للشيخ العجوز أن يكون ذو فائدة كبيرة للمجتمع من خلال تجربته الشخصية التي يمكنه تجسيدها في الأعمال الأدبية، أو السير الذاتية، أو الرسوم الفنية، أو الحديثات المثيرة. ربما ستكون تجربة كبار السن وحيوية الشبان أفضل توازن. لا أعتقد أن مثل هذا التوازن سيكون مستحيلاً. إنه يعتمد على الشيخ العجوز نفسه الذي يمكنه أن يتخيل الحياة كنبع للسرور والمتعة، ويستمتع إلى نداء الطبيعة في أغان لا تعد ولا تحصى، ويشعر أن وليمة الطبيعة هي وليمة الحب نفسه وأن الحياة بلا حب ستكون تافهة بالنسبة للشبان والكبار على حد سواء. ثم يمكن للشيخ العجوز أن يستثمر سنوات حياته في التمتع بالحياة. وبناءً على ذلك، يمكن أن تكون الشيخوخة مترادفة للشباب بطريقة تجعل الناس ينظرون إليها على أنها شكل من أشكال الحيوية التقدمية.

Table 1 Quantitative analysis:

	Source text	Human translation	GPT-3's translation
Sentences	8	5	8
Words	190	144	146
Average sentence length	24.63 words	29.60 words	19.13 words

The quantitative analysis revealed that the human translation used fewer sentences and words compared to GPT-3's translation. As shown in Table 1, the word count in the human translation is similar to that in the GPT-3 translation, and both translations contain fewer words than the source text. While GPT-3's translation maintained the same number of sentences as the source text but used fewer words, the human translation used fewer sentences than both the source text and GPT-3's translation. The average sentence length was higher in the human translation (29.6 words) compared to GPT-3's translation (19.13 words) because GPT-3 tends to use shorter sentences. Human translation adheres to the Arabic writing convention, which favors longer sentences with many commas (Alotaiby et al., 2009; Khafaji, 2001). Consequently, the human translation contained 5 sentences, whereas GPT-3's translation contained 8 sentences.

#### Accuracy:

Both translations effectively convey the main idea of the text, which discusses old age and its benefits to individuals and society. However, there are differences in word choice and phrasing that impact the accuracy of the translations. For instance, the human translation used the word (أحاديث) for "talks," while GPT-3 used (الحديثات), which does not exist in Arabic as an equivalent for "talks," as illustrated in Table 2. Despite these differences, both translations generally employ accurate and appropriate vocabulary to convey the original text's meaning.

Table 2. Examples from human translation vs. GPT-3's translation

The words/phrases	Human translation	GPT-3 translation
Talks	أحاديث	الحديثات

#### Fluency:

Both translations are grammatically correct, though ChatGPT makes two grammatical errors. First, the use of the pronoun of matter (ضمير الشأن) in (الشيخوخة هي مدرسة) introduces unnecessary redundancy, which the human translator avoided. Second, ChatGPT renders the predicate of the verb (يكون) as (ذو), while it should be in the accusative case (ذا). In contrast, the human translation omitted it entirely. Despite these issues, both translations are grammatically sound and easy to read. However, the human translation is more natural and fluid, employing more idiomatic expressions and phrasing. For example, the sentence "feel that the banquet of nature is the banquet of love itself and that life void of love would be trivial for the young and the old alike" is more naturally translated by the human translator as "وشعر "، وأن وليمة الطبيعة هي وليمة الحب ذاته، وأن الحياة الخالية من الحب تافهة بالنسبة إلى الشباب وإلى الشيوخ على حد سواء ويشعر أن وليمة الطبيعة هي وليمة الحب نفسه وأن الحياة بلا حب ستكون تافهة بالنسبة للشبان والكبار " whereas GPT-3's translation is "أن الحياة بلا " The phrase "الحياة الخالية من الحب تافهة" in the human translation is stronger than "الحياة بلا " used by GPT-3. On the other hand, GPT-3 translates "feel" into "يشعر" while the human translation uses "شعر". In Arabic, the present tense is used to indicate frequency and continuity, making GPT-3's translation more appropriate in this context. The results are detailed in Table 3.

Table 3. Examples from human translation vs. GPT-3's translation

The words/phrases	Human translation	GPT-3 translation
-------------------	-------------------	-------------------

feel that the banquet of nature is the banquet of love itself and that life void of love would be trivial for the young and the old alike	وشعر أن وليمة الطبيعة هي وليمة الحب ذاته، وأن الحياة الخالية من الحب تافهة بالنسبة إلى الشباب وإلى الشيوخ على حد سواء .	ويشعر أن وليمة الطبيعة هي وليمة الحب نفسه وأن الحياة بلا حب ستكون تافهة . بالنسبة للشباب والكبار على حد سواء .
Feel	شعر	يشعر

#### Cohesion and coherence

Regarding cohesiveness and coherence, the two Arabic translations exhibit differences in the use of cohesive devices. At the sentence level, the human translator's version extensively employs logical connectors, while GPT-3's translation completely omits these elements. For instance, the phrase "The truth" is translated by the human translator as (والحق), providing a logical connection to the previous sentence. In contrast, ChatGPT translates it as (الحقيقة), starting a new sentence without linking it to the previous one.

#### Translation of figurative language:

The sentence "old age is an expensive school because it is a storehouse of human experience" is translated by the human translator as فالشيخوخة مدرسة غالية لأنها مستودع للخبرة البشرية, while ChatGPT translates it as الشيخوخة هي مدرسة مكلفة لأنها مخزن لتجربة الإنسان (see Table 4). In this case, the human translation of "expensive school" as مدرسة غالية is more accurate than ChatGPT's مدرسة مكلفة. Another example is the translation of the simile "life as a spring of merriness and pleasure." The human translator uses a rhetorical simile, rendering it as الحياة ينبوع مرح ومسرة, whereas ChatGPT uses the simile marker (كـ), translating it as الحياة كنبع للسرور والمتعة. The human translation thus provides a stronger sense of the original text by employing a rhetorical simile rather than the straightforward simile used by ChatGPT. However, the phrase "numberless songs" is translated by the human translator as بأغان لا عدد لها, while GPT-3 translates it as أغان لا تعد ولا تحصى. In this case, GPT-3's translation appears to be more effective than the human translation.

Table 4. Examples from human translation vs. GPT-3's translation

The words/phrases	Human translation	GPT-3 translation
old age is an expensive school because it is a storehouse of human experience	فالشيخوخة مدرسة غالية لأنها مستودع للخبرة البشرية	الشيخوخة هي مدرسة مكلفة لأنها مخزن لتجربة الإنسان .
life as a spring of merriness and pleasure	الحياة ينبوع مرح ومسرة	الحياة كنبع للسرور والمتعة

#### Text 2

Source text	الأزهار أجمل الأشياء التي خلقها الله، فالعالم يسمي مكاناً موحشاً من دون أزهار تزينه؛ ويمسي وجهاً لا يبتسم أو وليمة لا يسمع فيها ترحاب المضيف. أليست الأزهار نجوم الأرض؟ وأليست النجوم أزهار السماء؟ الأزهار جواهر الطبيعة تستخدمها الطبيعة لتزيين نفسها، والحق أنه يصعب على المرء أن يحلل سحر الأزهار فذاك يشبه تشريح الموسيقى، وذاك أمر مستحيل، فالأفضل أن يستمتع المرء بالموسيقى بدلاً من فهمها، والشيء ذاته يقال في الأزهار الجميلة هي بسمات الملائكة، لذلك فإن المرء عندما يفلح حديقته فكأنه يمشي مع الملائكة، فهو يحدق في الأزهار ويرى في كل ورقة من وريقاتها كتاباً سطرته الملائكة تسبيحاً بحمد الله وتمجيدياً لعظمته، فأقل الأزهار شأناً ترفع رأسها وتنظر بمسرة معبرة بصورة سرية عن عظمة خالقها، لذلك يولي الإنسان أهمية للأزهار فيقدمها في عديد من المناسبات، عند الولادة والزواج والولائم الخ.....
Human translation	Flowers are the most beautiful things that God created. The world would be a desolate place without flowers. It would become an unsmiling face or a banquet where no welcome is heard. Aren't flowers the stars of the earth? Aren't stars the flowers of heaven? Flowers are jewels with which nature beautifies herself. In actuality, it would be difficult for one to analyze the charm of flowers. It would be like dissecting music, which is impossible. Man had better enjoy music rather than understand it. The same thing can be said of flowers. Beautiful flowers are the smiles of angels. Therefore, when one is cultivating one's garden it is as if one were walking with angels. One stares at flowers and sees in every leaf a book written by angels in praise of God's greatness. The least flower raises up its head, looking with pleasure and expressing secretly the greatness of its creator. It is for this reason that man takes care of flowers and presents them on numerous occasions, at birth, marriage and at banquets etc.
GPT-3's translation	Flowers are among the most beautiful things created by God. The world becomes a desolate place without the adornment of flowers. It becomes a face that doesn't smile, or a feast without the host's hospitality. Aren't flowers the stars of the earth? And aren't the stars the flowers of the sky? Flowers are nature's gems used to embellish itself. It's indeed difficult for anyone to dissect the magic of flowers; it's akin to dissecting music, something close to impossible. It's better to enjoy music rather than trying to understand it. The same can be said about flowers.

Beautiful flowers are the smiles of angels. When one tends to a garden successfully, it's like walking with the angels. You gaze at the flowers and see in each petal a book written by the angels, praising and glorifying God's greatness. Even the humblest of flowers lift their heads and express, in their own secret way, the magnificence of their Creator. That's why humans attach significance to flowers and present them on various occasions, such as births, weddings, and festive gatherings, and so on.

*Table 5 Quantitative analysis:*

	Source text	Human translation	ChatGPT's translation
Sentences	3	15	14
Words	130	177	181
Average sentence length	44 words	12.67 words	13.79 words

The quantitative analysis reveals that the human translation used fewer sentences and words. Table 5 shows that the word count in the human translation is similar to that in the ChatGPT translation, with both translations containing more words than the source text. Both translations simplify the source text by breaking it into shorter sentences. While the source text contains 3 sentences, the human translation has 15 sentences, and the ChatGPT translation has 14. The average sentence length is also comparable in both translations, with the human translation averaging 12.67 words per sentence and the ChatGPT translation averaging 13.79 words per sentence.

*Accuracy:*

Both translations convey the main idea of the text, which discusses flowers and their significance to humans. However, there are differences in word choice and phrasing that impact the accuracy of the translations. For instance, the word *فهر* is translated as "one" in the human translation, a neutral pronoun suitable for different genders. Conversely, GPT-3 uses a rhetorical device called "enallage" (الالتفات), changing the third person pronoun to a second person pronoun. Regarding word choice, both translations use accurate and appropriate vocabulary to convey the original text's meaning. For example, both translations use "a desolate place" for the Arabic phrase "مكاناً موحشاً". However, there are some differences in word choice. In the human translation, "charm" is used for the word *سحر*, while GPT-3 translates it as "magic." In this context, "charm" is more accurate than "magic." Additionally, the word *نفسها* is incorrectly translated as "herself" in the human translation, while GPT-3 correctly translates it as "itself." The results are shown in Table 6.

*Table 6. Examples from human translation vs. GPT-3's translation*

The words/phrases	Human translation	GPT-3 translation
سحر	Charm	Magic
مكاناً موحشاً	a desolate place	a desolate place

*Fluency:*

Both translations are grammatically correct and easy to understand. However, the human translation is more natural and flows better, employing idiomatic expressions and phrasing effectively. For instance, the phrase "فإن المرء عندما يفلح حديقته فكأنه يمشي مع الملائكة" is rendered by the human translator as "when one is cultivating one's garden it is as if one were walking with angels," which feels more natural compared to GPT-3's translation, "When one tends to a garden successfully, it is like walking with the angels." In GPT-3's translation, the inclusion of "successfully" is unnecessary in this context and not present in the source text.

*Table 7. Examples from human translation vs. GPT-3's translation*

The words/phrases	Human translation	GPT-3 translation
فإن المرء عندما يفلح حديقته فكأنه يمشي مع الملائكة	when one is cultivating one's garden it is as if one were walking with angels	When one tends to a garden successfully, it is like walking with the angels

*Cohesion and coherence*

Both translations are coherent and well-structured. However, the human translation exhibits greater cohesion by employing more connectives and transitional phrases that effectively link ideas and sentences together. For instance, the use of "Therefore" in the human translation establishes a logical connection to the preceding sentence, whereas ChatGPT's translation omits this word.

*Translation of figurative language:*

The sentence "الأزهار جواهر الطبيعة تستخدمها الطبيعة لتزين نفسها" is translated by the human translator as "Flowers are jewels with which nature beautifies herself," whereas ChatGPT translates it as "Flowers are nature's gems used to embellish itself." The human translation employs more precise terminology compared to ChatGPT's translation. These findings are detailed in Table 8.

*Table 8. Examples from human translation vs. GPT-3's translation*

The words/phrases	Human translation	GPT-3 translation
-------------------	-------------------	-------------------

الأزهار جواهر الطبيعة تستخدمها الطبيعة لتزيين نفسها	Flowers are jewels with which nature beautifies herself	Flowers are nature's gems used to embellish itself
---	---	--

### Text 3

Source text	Why We Should Be Generous Generosity is a likeable, praiseworthy quality. In actuality, this quality brings out several other qualities. To be generous would be to be magnanimous and to be magnanimous would be to be courageous. Magnanimity and courage are both a kind of remarkable manliness. Probably we can assess generosity correctly if we contemplate its opposite, namely, meanness. Thus, we could determine whether it would be in one's interest to be generous. Generosity implies open-mindedness. A generous person is one who not only spends freely but also looks at life without fanaticism or narrow-mindedness. He is a giver intellectually and materially.	
Human translation	لماذا يجب أن نكون كرماء؟ الكرم خصلة من الخصال الحميدة التي يجبها المرء في الناس. الحق أن هذه الصفة تبرز بضع صفات أخرى، فإن يكون المرء كريماً فذلك يعني أن يكون شهماً، وأن يكون شهماً فذلك يعني أن يكون شجاعاً، فالشجاعة والشجاعة ضربان من ضروب الرجولة الفذة. لعنا نستطيع أن نقدر الكرم على نحو صحيح إذا تأملنا نقيضه نعني بذاك الشح، فالنقيض يمكننا أن نقرر إذا كان في صالح المرء أن يكون كريماً، والكرم ينطوي على انفتاح فكري، فالكرم ليس من ينفق بسخاء فحسب بل إنه رجل يقدر على أن ينظر إلى الحياة من دون تعصب أو ضيق فكري. إنه معطاء من الناحية الفكرية والمادية.	
GPT-3's translation	لماذا يجب أن نكون سخاء السخاء هو صفة محبوبة ومحمودة في الواقع، تبرز هذه الصفة عدة صفات أخرى. أن تكون سخياً يعني أن تكون كريماً، وأن تكون كريماً يعني أن تكون شجاعاً. الكرم والشجاعة هما نوع من أنواع الرجولة الملحوظة. ربما يمكننا تقييم السخاء بشكل صحيح إذا فكرنا في النقيض منه، أي البخل. وبالتالي، يمكننا تحديد ما إذا كان في مصلحة الشخص أن يكون سخياً. السخاء يعني التسامح. الشخص السخي هو الذي لا ينفق بسخاء فقط ولكنه أيضاً ينظر إلى الحياة بدون تعصب أو ضيق أفق. إنه من يقدم بفكره ومواده.	

Table 9 Quantitative analysis:

	Source text	Human translation	ChatGPT's translation
Sentences	9	5	9
Words	105	105	91
Average sentence length	12.22 words	21.60 words	11.00 words

Table 9 indicates that the human translation used fewer sentences and fewer words. Human translation adheres to the Arabic paragraph structure, characterized by longer sentences punctuated with commas. As a result, the human translation comprised 5 sentences compared to 9 in ChatGPT's translation. While the number of words in the human translation closely matched that of the source text, ChatGPT used fewer words overall. Despite maintaining the same number of sentences, ChatGPT's translation contained fewer words. The average sentence length was higher in the human translation (21.60 words) and lower in ChatGPT's translation (11.00 words), reflecting ChatGPT's tendency to use shorter sentences.

#### Accuracy:

Both translations effectively convey the main idea of the text, which centers around generosity. However, there are differences in word choice and phrasing that impact the accuracy of the translations. For instance, the human translation uses the word "كرم" for "generous," while ChatGPT opts for "السخاء," which is a synonym. Another instance is the translation of "open-mindedness," where ChatGPT translates it as "التسامح," while the human translation accurately renders it as "انفتاح فكري," which better captures the intended meaning. Similarly, for the word "meanness," ChatGPT translates it as "البخل," whereas the human translation uses "الشح," which conveys a stronger sense. This highlights the human translator's ability to select words with different meanings. Additionally, the translation of "magnanimous" is accurately rendered as "شهماً" by the human translation, whereas ChatGPT translates it incorrectly as "كريماً." These findings are summarized in Table 10.

Table 10. Examples from human translation vs. GPT-3's translation

The words/phrases	Human translation	GPT-3 translation
Generous	كرم	السخاء
open-mindedness	انفتاح فكري	التسامح
Meanness	الشح	البخل
Magnanimous	شهماً	كريماً

#### Fluency:

Both translations are grammatically correct and clear, yet the human translation stands out for its naturalness. It incorporates more idiomatic expressions and fluid phrasing. For instance, the sentence "Magnanimity and courage are both a kind of remarkable manliness" is translated in the human version as "فالشجاعة والشجاعة ضربان من ضروب الرجولة الفذة" which flows more naturally compared to ChatGPT's



translation, "الكرم والشجاعة هما نوع من أنواع الرجولة الملحوظة". Furthermore, the use of the word "both" in this context introduces unnecessary redundancy. Additionally, the singular form of the word "نوع" in ChatGPT's translation is morphologically incorrect, as the pronoun of matter indicates a dual form, necessitating "نوعان" instead of "نوع". These findings are summarized in Table 11.

Table 11. Examples from human translation vs. GPT-3's translation

The words/phrases	Human translation	GPT-3 translation
Magnanimity and courage are both a kind of remarkable manliness	فالشجاعة والشجاعة ضربان من ضرب الرجولة الفذة.	الكرم والشجاعة هما نوع من أنواع الرجولة الملحوظة.
Both	ضربان	نوع

#### Cohesion and coherence

Both translations are coherent and well-structured. However, the ChatGPT uses more connectives and transitional phrases that help to link the different ideas and sentences together. For example, it uses phrases like "وبالتالي" and "في الواقع".

#### Translation of figurative language:

He is a giver intellectually and materially" is translated by human as "إنه معطاء من الناحية الفكرية والمادية". While it is translated by ChatGPT as "إنه من يقدم بفكره ومواده". It is more accurate in human translation than in ChatGPT's translation. The results are shown in Table 12.

Table 12. Examples from human translation vs. GPT-3's translation

The words/phrases	Human translation	GPT-3 translation
He is a giver intellectually and materially	إنه معطاء من الناحية الفكرية والمادية	إنه من يقدم بفكره ومواده

#### Text 4

Source text	لا بد لنا عند الحديث على خصال الجار الطيب أن نأخذ بعين الاعتبار بأن الجار امرؤ يعيش على مقربة منا. إن مفهوم القرب ينطوي على إقامة علاقات طيبة، والمفهوم ذاته يعني التواصل المستمر. بالمعنى الاجتماعي الجار هو شخص يعيش إلى جوارك، وأحسب أن ذلك بالضرورة اعتبار الجار واحداً من أفراد العائلة، لك يعني هذا الاعتبار يجب أن يكون مبنياً على صفات معينة لا بد أن يتحلى الجار بها. فالجار يجب أن يكون شريفاً على نحو يمكن أن يعتبر واحداً من أفراد العائلة، إن هذا الاعتبار يقوده إلى احترام حرمة جاره، فتصرفه الفردي ومعاملته يجب أن يكونا بصورة جوهرية مؤسسين على عدم تجاوز الحدود..
Human translation	When talking about the qualities of a good neighbour we should realize that a neighbour is a person who lives next door. The concept of nearness implies the establishing of good relations. The same concept means constant communication. In the social sense, a neighbor is a person who lives close to you and that would necessarily involve, I think, considering a neighbor as if he were a family member. However, this consideration must be founded on certain qualities that a neighbor should possess. A neighbor should be honest if he is to be considered a family member. This consideration should motivate him to respect the sanctity of his neighbor. His individual behavior and treatment should be essentially founded on non-transgression. In other words, his honor and that of his neighbor should be the same.
GPT-3's translation	When talking about the qualities of a good neighbor, we must consider that a neighbor is someone who lives in close proximity to us. The concept of proximity involves building good relationships, and this concept itself means continuous communication. Socially, a neighbor is a person who lives next to you, and I believe that necessarily considering the neighbor as one of the family members means that this consideration should be based on certain qualities that the neighbor must possess. The neighbor should be honorable in a way that allows them to be considered as one of the family members. This consideration leads to respecting the sanctity of the neighbor. Their individual behavior and treatment should fundamentally be established on not crossing boundaries.

Table 13 Quantitative analysis:

	Source text	Human translation	ChatGPT's translation
Sentences	4	10	6
Words	103	135	122
Average sentence length	26.50 words	14.30 words	21.17 words

Table 13 shows that the quantitative analysis, the human translation used more sentences (10 sentences) and more words (135 words). The number of words in both translations is more than the words in the source text. The average sentence length was low in the human translation (14.30 words) and high in the ChatGPT's translation (21.17 words).

#### Accuracy:

Both translations effectively convey the main idea of the text, which concerns neighbors and their rights. However, there are differences in word choice and phrasing that affect the translations' accuracy.

For instance, the human translation uses the word "person" for "امرؤ," whereas ChatGPT translates it as "someone," which is correct but less precise than "person." Another example is the translation of "شريفاً." The human translation uses "honest," while ChatGPT opts for "honorable," which is more accurate. The word "حرمة" is translated as "sanctity" in both versions. Additionally, the pronoun "فتصرفه الفردي" is correctly translated as "his" in the human translation but inaccurately rendered as "their" in ChatGPT's translation. The results are detailed in Table 14.

Table 14. Examples from human translation vs. GPT-3's translation

The words/phrases	Human translation	GPT-3 translation
امرؤ	Person	Someone
شريفاً	Honest	Honorable
حرمة	Sanctity	sanctity

#### Fluency:

Both translations are grammatically correct and coherent, yet the human translation is more fluent and natural, employing idiomatic expressions and smoother phrasing. For instance, the phrase "عدم تجاوز الحدود" is translated by ChatGPT as "crossing boundaries," which is more natural than the human translation's "non-transgression." Additionally, the human translation enhances clarity by adding "In other words, his honor and that of his neighbor should be the same" to explain this concept. These enhancements contribute to the human translation's clarity and readability. The comparative results are presented in Table 15.

Table 15. Examples from human translation vs. GPT-3's translation

The words/phrases	Human translation	GPT-3 translation
عدم تجاوز الحدود	non-transgression	crossing boundaries

#### Cohesion and coherence

Both translations demonstrate coherence and organization; however, the human translation exhibits greater cohesion by employing more connectives and transitional phrases that effectively link ideas and sentences together. For instance, the sentence "بالمعنى الاجتماعي الجار هو شخص يعيش إلى جوارك، وأحسب أن ذلك بالضرورة اعتبار الجار واحدا من أفراد العائلة، لك يعني هذا الاعتبار يجب أن يكون مبنياً على صفات معينة لا بد أن يتحلى الجار بها" is translated in the human version as, "In the social sense, a neighbor is a person who lives close to you, and that would necessarily involve, I think, considering a neighbor as if he were a family member. However, this consideration must be founded on certain qualities that a neighbor should possess." In contrast, ChatGPT translates the same sentence as, "Socially, a neighbour is a person who lives next to you, and I believe that necessarily considering the neighbour as one of the family members means that this consideration should be based on certain qualities that the neighbor must possess." ChatGPT retains the sentence structure from the source text without introducing a linking word like "However." Moreover, the human translation includes the phrase "I think" as a parenthetical expression, whereas ChatGPT integrates it into the main sentence. These distinctions illustrate how the human translation uses explicit connectors and structural adjustments to enhance coherence, while ChatGPT maintains a more literal rendering of the original text. Detailed comparisons are provided in Table 16.

Table 16. Examples from human translation vs. GPT-3's translation

The words/phrases	Human translation	GPT-3 translation
For example, the sentence "بالمعنى الاجتماعي الجار هو شخص يعيش إلى جوارك، وأحسب أن ذلك بالضرورة اعتبار الجار واحدا من أفراد العائلة، لك يعني هذا الاعتبار يجب أن يكون مبنياً على صفات معينة لا بد أن يتحلى الجار بها"	In the social sense, a neighbor is a person who lives close to you and that would necessarily involve, I think, considering a neighbor as if he were a family member. However, this consideration must be founded on certain qualities that a neighbor should possess.	Socially, a neighbour is a person who lives next to you, and I believe that necessarily considering the neighbour as one of the family members means that this consideration should be based on certain qualities that the neighbor must possess.

#### Translation of figurative language:

The sentence "الجار امرؤ يعيش على مقربة منا" is rendered in the human translation as "a neighbour is a person who lives next door," whereas in ChatGPT's translation it is presented as "a neighbour is someone who lives in close proximity to us." Semantically, the human translation is more precise than ChatGPT's translation. The term "close proximity" in ChatGPT's rendering could imply various types of proximity beyond physical proximity, such as proximity in work or ideology. Detailed results are outlined in Table 17.

Table 17. Examples from human translation vs. GPT-3's translation

The words/phrases	Human translation	GPT-3 translation
-------------------	-------------------	-------------------

الجار امرؤ يعيش على مقربة منا	a neighbour is a person who lives next door	a neighbour is someone who lives in close proximity to us
-------------------------------	---	---

## Discussion

This study seeks to explore the efficacy of ChatGPT in translating figurative language relative to human translation. The primary research question centers on assessing how effectively ChatGPT translates figurative language compared to human translators.

The findings from the analysis of the four texts used in this study demonstrate that when translating English texts into Arabic, GPT-3 attempts to preserve the original text's sentence count. However, when the source text (ST) is in Arabic, GPT-3 tends to shorten lengthy Arabic sentences by dividing them into shorter segments. This adaptation is influenced by the structural difference between English and Arabic sentences (Berrichi & Mazroui, 2021). In contrast, human translation prioritizes semantic fidelity over structural fidelity. While ChatGPT strives to uphold both meaning and structure, human translators focus primarily on conveying the intended meaning. This approach leads to variations in sentence structure and length between translations from English to Arabic and vice versa.

Human translators do not restrict themselves to specific word counts per sentence, whereas ChatGPT aims to minimize average sentence length. Consequently, this difference in focus also impacts the average sentence length observed in the translations.

### *Accuracy*

In all examined texts, both translations effectively convey the central ideas. Nonetheless, differences in word choice and phrasing occasionally impact the accuracy of the translations. The human translation often employs more precise terminology than the ChatGPT translation. Conversely, there are instances where ChatGPT selects more accurate terms than the human translation, exemplified by its translation of "الإنسان" as "humans" versus the human translation's "man". Furthermore, human translation excels in handling synonymous words, while ChatGPT sometimes struggles with pronoun translations. This issue is evident in a few instances of human translation, such as the incorrect rendering of "نفسها" as "herself" instead of "itself".

### *Fluency:*

Both translations maintain grammatical accuracy and readability. However, the human translation stands out for its naturalness and smooth flow. It achieves this by integrating a greater number of idiomatic expressions and different phrasing, which enrich the linguistic quality and authenticity of the text. In contrast, while the other translation remains grammatically correct and clear, it may not capture the distinctive linguistic differences that contribute to the naturalness and fluidity found in the human-produced version.

### *Cohesion and coherence*

Both translations exhibit coherence and well-structured composition. In some cases, the human translation shows enhanced cohesion through the use of additional connectives and transitional phrases that effectively connect ideas and sentences. Conversely, ChatGPT's translation also demonstrates strong cohesion by employing a variety of connectives and transitional phrases, indicating its capability in maintaining coherence in translated texts.

### *Translation of figurative language:*

In many instances of figurative language, human translation proves to be more accurate than ChatGPT's rendition. This underscores humans' proficiency in capturing the differences and complexities inherent in figurative expressions. However, there are specific instances, such as in the case of similes, where differences arise. For example, human translators may omit the word "as" and translate directly, which could reflect stylistic choices or linguistic norms in the source language. In contrast, ChatGPT often retains "as" and uses the Arabic equivalent "كـ" in its translations. Despite the general tendency for human translation to excel in handling figurative language, there are instances where ChatGPT performs better. This study aligns with Banat and Adla's findings (2023), suggesting that GPT-3 demonstrates reasonable accuracy and fluency in translating religious texts.

Machine translation models like ChatGPT may struggle with accurately and contextually translating figurative language. Figurative expressions often hinge on cultural differences, context, and subtle meanings that pose challenges for these models. While ChatGPT can comprehend and generate human-like text, it may not consistently capture the intended figurative differences with the depth and precision of human translators. Human translators bring cultural and contextual understanding, creativity, and intuition to the translation process. They adeptly navigate idioms, metaphors, and other figurative elements, ensuring that the translated text preserves the intended meaning and emotional resonance. While machine translation, including ChatGPT, continues to evolve and improve, there remain limitations in fully capturing the richness of figurative language. For the most accurate and

culturally sensitive translations, especially involving figurative language, human translators remain indispensable. Their linguistic expertise, cultural insights, and different grasp of context enable them to produce translations that surpass current machine translation capabilities.

## Conclusion

The comparative study between human translators and ChatGPT in translating figurative language from Arabic to English offers valuable insights into the capabilities and limitations of AI-driven translation technologies. While ChatGPT demonstrates competence in generating generally comprehensible translations, it struggles with the differences inherent in metaphorical language. In contrast, human translations exhibit higher levels of naturalness, fluency, and coherence, crucial for capturing the essence of the original text. Quantitative analysis underscores methodological distinctions: human translations tend to elaborate more with a greater number of words and sentences, potentially enhancing readability and contextual richness. However, this can introduce unnecessary complexity. ChatGPT translations, characterized by shorter sentences and a more direct style, may occasionally omit connectives and transitional phrases that contribute to text fluidity.

Despite these differences, both human and ChatGPT translations preserve the fundamental themes of the source texts. ChatGPT's translations have even surpassed human efforts in specific instances, highlighting the potential value of AI translation tools in certain contexts. Nevertheless, the study underscores that human translators currently excel in contexts requiring deep cultural awareness, familiarity with colloquial idioms, and different understanding of figurative language.

The implications of this study extend beyond a mere comparison of translation systems. They underscore the potential for a hybrid approach that leverages both human expertise and AI efficiency. As AI technology advances, there is promise in a collaborative model where AI handles routine translation tasks, allowing human translators to focus on refining and enhancing translations with their profound linguistic and cultural insights.

### *Limitations*

The study has some limitations. First, the study may not cover the full range of figurative language, potentially overlooking certain differences and idiomatic expressions unique to Arabic that could affect the results. Second, the study might not fully account for the deep cultural context and subtleties that human translator can inherently understand and incorporate, which is crucial for accurate translation of figurative language. Third, focusing on quantitative measures, such as word and sentence count, may overlook qualitative aspects of translation, such as emotional tone, stylistic choices, and the subtleties of literary language. Fourth, the study analysed four texts. Future studies can translate more texts with a variety of figurative language used. Finally, the study's findings are dependent on the current state of AI technology, specifically ChatGPT, and may not reflect future advancements that could significantly improve AI translation capabilities.

### *Implications from the study*

The study suggests a promising future for hybrid translation models that enhance both AI efficiency and human expertise and improve the overall quality and accuracy of translations. In addition, AI can be used to handle simpler translation tasks, thereby freeing human translator to concentrate on more complex translations that require a deep understanding of cultural differences, idiomatic expressions, and figurative language. Moreover, the areas where ChatGPT struggles, particularly with figurative language and cultural context, highlight specific areas for AI developers to focus on for future improvements. The study provides valuable insights for educational programs in translation studies, emphasizing the importance of training human translator in the differences of language that AI currently cannot master. Organizations can strategically use AI translation tools for efficiency in basic translation tasks while relying on human translator for more different tasks, optimizing resources and improving translation quality overall.

## References

- Abdel-Raheem, Ahmed. 2024. The creative minds of Arab cartoonists: metaphor, culture and context. *Text & Talk*, 442, 141-174. <https://doi.org/10.1515/text-2021-0100>
- Abdul Wahid, Puteri Roslina. 2017. Translation of figurative language from English to Malay: An Analysis of the Translation of the Harry Potter Series. In *Proceedings of Research World International Conference, Phuket, Thailand, 9th-10th February*.
- Aghai, Mohammed. 2024. ChatGPT vs. Google Translate: Comparative Analysis of Translation Quality. *Iranian Journal of Translation Studies*, 2285, 87-103.
- Al-Hasan, Tamim Mahmud, Sayed, Aya Nabil, Bensaali, Faycal, Himeur, Yassine, Varlamis, Iraklis and Dimitrakopoulos, George. 2024. From Traditional Recommender Systems to GPT-Based

- Chatbots: A Survey of Recent Developments and Future Directions. *Big Data and Cognitive Computing*, 84, 1-28. <https://doi.org/10.3390/bdcc8040036>
- Alotaiby, Fahad, Ibrahim Alkharashi, and Salah Foda. 2009. Processing large Arabic text corpora: Preliminary analysis and results. In *Proceedings of the second international conference on Arabic language resources and tools* pp. 78-82.
- Baker, Mona. 2018. *In other words: A coursebook on translation*. Routledge.
- Banat, Maysaa, and Yasmine Abu Adla. 2023. Exploring the Effectiveness of GPT-3 in Translating Specialized Religious Text from Arabic to English: A Comparative Study with Human Translation. *Journal of Translation and Language Studies*, 42, 1-23.
- Bell, Roger T., and Christopher N. Candlin. 2016. *Translation and translating: Theory and practice*. London and New York: Routledge.
- Berrichi, Safae, and Azzeddine Mazroui. 2021. Addressing limited vocabulary and long sentences constraints in English–Arabic neural machine translation. *Arabian Journal for Science and Engineering*, 469, 8245-8259.
- Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in Neural Information Processing Systems*, 29.
- Cai, Zhenguang, Xufeng Duan, David A. Haslett, Shuqi Wang, Martin J. Pickering. 2023. Does ChatGPT resemble humans in language use?. *arXiv preprint arXiv:2303.08014*.
- Cingillioglu, Ilker. 2023. Detecting AI-generated essays: the ChatGPT challenge. *The International Journal of Information and Learning Technology*, 403, 259-268. <https://doi.org/10.1108/IJILT-03-2023-0043>
- Crompton, Helen, and Diane Burke. 2023. Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education*, 201, 1-22. <https://doi.org/10.1186/s41239-023-00392-8>
- Eke, Damian Okaibedi. 2023. ChatGPT and the rise of generative AI: threat to academic integrity?. *Journal of Responsible Technology*, 13, 100060.
- Gao, Catherine, Frederick M. Howard, Nikolay S. Markov, Emma C. Dyer, Siddhi Ramesh, Yuan Luo & Alexander T. Pearson. 2022. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *BioRxiv*, 2012–2022.
- Haman, Michael, and Milan Školník. 2023. Using ChatGPT to conduct a literature review. *Accountability in Research*, 1–3.
- Hendy, Amr, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify, and Hany Hassan Awadalla. 2023. How good are GPT models at machine translation? a comprehensive evaluation. *arXiv preprint arXiv:2302.09210*.
- Javaid, Mohd, Abid Haleem, Ravi Pratap Singh, Shahbaz Khan, Ibrahim Haleem Khan M., Haleem. 2023. Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 32, 100115. <https://doi.org/10.1016/j.tbench.2023.100115>
- Jiao, Wenxiang, Wenxuan Wang, Jen-tse Huang, Xing Wang, Shuming Shi, Zhaopeng Tu. 2023. Is ChatGPT a good translator? A preliminary study. *arXiv preprint arXiv:2301.08745*.
- Khafaji, Rasoul. 2001. Punctuation marks in original Arabic texts. *Zeitschrift für arabische Linguistik*, 40, 7-24.
- Khoshafah, Faten. 2023. ChatGPT for Arabic-English translation: Evaluating the accuracy. Research Square. <https://doi.org/10.21203/rs-2814154/v1>.
- Lee, Tong King. 2023. Artificial intelligence and posthumanist translation: ChatGPT versus the translator. *Applied Linguistics Review*. <https://doi.org/10.1515/applirev-2023-0122>
- Lund, Brady, and Ting Wang. 2023. Chatting about ChatGPT: how may AI and GPT impact academia and libraries? *Library Hi Tech News*. 403, 26-29 <https://doi.org/10.1108/LHTN-01-2023-0009>
- Luo, Renqian, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. 2022. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics*, 236, 1-11. <https://doi.org/10.1093/bib/bbac409>
- Montgomery, Martin, Durant, Alan, Fabb, Nigel, Furniss, Tom, and Mills, Sara. 2007. *Ways of reading: Advanced reading skills for students of English literature*. Routledge.
- Nida, Eugene. 1975. *Language Structure and Translation*. California, United States: Stanford University Press.
- Qiong, Liu., and Zhang Xiaobing. 2005. Towards the translation of figurative language. *Canadian Social Science*, 11, 122- 126. <http://dx.doi.org/10.3968/j.css.1923669720050101.022>



## تقييم ترجمة اللغة المجازية: دراسة مقارنة بين ChatGPT والمترجمين البشر

### المستخلص

في الأونة الأخيرة أظهرت أداة الذكاء الاصطناعي (ChatGPT) تقدماً كبيراً في عدد من المهام المتعلقة باللغة، بما في ذلك الترجمة الآلية. ومع ذلك، تركزت العديد من الدراسات التي قامت بتقييم أداء ChatGPT في مهام الترجمة على النصوص العامة. لذا كان الهدف الرئيسي لهذه الدراسة تقييم مدى قدرة GPT-3 على ترجمة محتوى يتضمن لغة مجازية من العربية إلى الإنجليزية والعكس، ومقارنة أدائه مع أداء المترجمين. ولتحقيق هذا الهدف، استخدمت الدراسة بعض النصوص التي تركز على مواضيع مختلفة تتضمن استخدام اللغة المجازية. خضعت هذه النصوص للترجمة من قبل مترجم بشري محترف وGPT-3. وقامت الدراسة بتقييم أداء ترجمة GPT-3 مقابل الترجمة البشرية باستخدام معايير نوعية. وشملت المعايير الدقة، والسلاسة، والتماسك، والترجمة للغة المجازية. أظهرت نتائج تحليل هذه المقاطع أن GPT-3 قام بترجمة مفهومة بشكل عام، ولكنها أخفقت في التنبيه للغة المجازية بالمقارنة مع الترجمة البشرية. كما أشارت النتائج إلى أن GPT-3 استخدم نفس العدد من الجمل في النص العربي عندما تم ترجمته من اللغة الإنجليزية. ومع ذلك عندما كان النص الأصلي باللغة العربية، فقد قام GPT-3 بتقليص طول الجمل العربية الطويلة وقسمها لتكون أقصر. تسلطت الدراسة الضوء على الإمكانيات المتاحة لنماذج الترجمة الهجينة التي تجمع بين كفاءة الذكاء الاصطناعي وخبرة الإنسان، مما يوحي بأن الذكاء الاصطناعي يمكنه التعامل مع المهام البسيطة بينما يترك الترجمة البشرية للترجمات المعقدة التي تتطلب فهماً للثقافة والتعابير اللغوية. كما تؤكد الدراسة على ضرورة تحسين الذكاء الاصطناعي في مجال اللغة المجازية والسياق الثقافي، وتقديم إرشادات للبرامج التعليمية لتدريب المترجمين البشر، وتوصي باستخدام الذكاء الاصطناعي بشكل استراتيجي للمهام الأساسية لتحسين استخدام الموارد وتحسين جودة الترجمة بشكل عام.

الكلمات الرئيسية: الترجمة بناءً على الذكاء الاصطناعي، ChatGPT، اللغة المجازية، الترجمة البشرية، الترجمة